DTIC FILE COPY

(2)

BBN SYSTEMS AND TECHNOLOGIES

AD-A228 993

27 September 1990

Office of Naval Research
Code 1133, Annual Report
800 N. Quincy Street
Arlington, VA 22217-5000

DTIC
ELECTE
OCT 15 1990
S E D

Dear Dr. van Tilborg:

Enclosed is the annual report requested for the contract titled Development of a Spoken Language System, contract #N00014-89-C-0008.

Sincerely,

Madeleine Bates

Madeleine Bates, Ph.D.
Assistant Department Manager
Speech and Natural Language Department

MB/hg

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

A Division of Bolt Beranek and Newman Inc.
10 Moulton Street Cambridge, Massachusetts 02138 Telephone: 617 873-3000

ANNUAL REPORT

Principal Investigator: John Makhoul
PI Institution: BBN Systems and Technologies
PI Phone Number: 617-873-3332
PI E-mail Address: makhoul@bbn.com
Grant or Contract Title: Development of a Spoken Language System
Grant or Contract Number: N00014-89-C-0008
Reporting Period: 1 Oct 89 - 30 Sep 90

1. Productivity Measures

Refereed papers submitted but not yet published: 0

Refereed papers published: 16

Unrefereed reports and articles: 1

Books or parts thereof submitted but not yet published: 1

Books or parts thereof published: 1

Patents filed but not yet granted: 0

Patents granted: 0

Invited presentations: 3

Contributed presentations: 0

Honors received:
Participation in various SLS committees
Co-chair 3rd ACL Conference on Applied NL Processing (1992)

Prizes or awards received: 0

Promotions obtained: 0

Graduate students supported : 0

Post-docs supported: 0

Minorities supported: 0

Principal Investigator: John Makhoul
PI Institution: BBN Systems and Technologies
PI Phone Number: 617-873-3332
PI E-mail Address: makhoul@bbn.com
Grant or Contract Title: Spoken Language Systems
Grant or Contract Number: N00014-89-C-0008
Reporting Period: 1 Oct 89 - 30 Sep 90

2. Detailed summary of technical progress

The goal of this research project is to integrate speech and natural language technologies into a spoken language system capable of understanding and responding to spoken English for interactive human-machine military applications, such as command and control, training of military personel, and logistics planning. The system we are building, called HARC (Hear And Respond to Continuous speech) will include a capability to adapt to new speakers and a capability to detect when a user says a new word.

Major accomplishments this year include:

1) top performance in the first multi-site competitive evaluation;

2) speedup of the natural language ~~processing~~ processing component (Delphi);

3) rapid port to the ATIS domain;

4) extending the discourse mechanism, including the handling of more general forms of definite reference;

5) selection of the TRANSCOM domain for our demonstration SLS system;

6) integration of speech recognition and natural language processing;

7) detecting and adding new vocabulary words in speech;

8) several new more efficient search algorithms for speech processing; AND

9) demonstration of real-time speech recognition with N-Best sentence output.

Common Evaluation on ATIS Domain

In the common evaluation performed in June, 1990, our natural language systems had the best overall natural language understanding performance. Out of 90 Class A test queries, the Delphi NL system produced 52 correct answers, 0 incorrect answers, and 38 no answers. The number of correct responses was one of the highest of all the systms tested, and the number of incorrect responses was by far the lowest. The Parlance NL system running on the same test data produced 58 correct answers, 6 incorrect answers, and 26 no answers. This number of correct answers was the highest of the systems tested.

Data Collection and Evaluation Methodology

We implemented and demonstrated a "Wizard scenario" for data collection, in which speech is elicited from a subject by presenting them with a computer system that appears to understand them (that is, the computer presents responses to what the subject says), but which actually has a human "wizard" listening to the spoken questions and typing the commands to the system to produce the answer. This methodology was adopted by TI for use in collecting the development and test data that was used in the formal system evaluation.

We continued our strong participation in developing a methodology for common evaluation of spoken language spstems, especially the evaluation of NL understanding systems. The proposals we made originally a year ago were finally adopted by the DARPA community with little modification. We advocated the use of objective evaluation based on canonical answers, we defined a format for canonical answers, and we wrote and distributed comparator software to other SLS sites to be used to compare their systems' answers with the canonical answers. We participated fully in various committees established to create the evaluation methodology, including helping TI to put together the relational ATIS database and to collect data via a Wizard data collection scenario.

Speech / NL Integration

Evaluation of N-Best Search Strategy

During the previous year we proposed a major new paradigm for integrating speech recognition and natural language, called the N-Best Paradigm. The basic idea was to use acoustic and statistical language models to find the N most likely whole sentence hypotheses, and then to pass these scored text strings on to the natural language component, which further filters and rescores the strings. The result is an extremely simple and efficient method for integrating speech and natural language. We also developed an efficient algorithm that would find the N-Best sentences. Since announcing this new strategy in October, 1989 at the DARPA workshop, most of the other research sites have adopted the N-Best Paradigm in their SLS systems.

During this year we performed many experiments with the N-Best paradigm. In particular, we found that when we used a fully connected statistical grammar with perplexity 100, the correct sentence was included within the 100 best sentence hypotheses 96% of the time! In all cases, there was at least one sentence within the list that would be perfectly acceptable to the natural language components. Thus, this paradigm would cause no search errors.

Approximate N-Best Search Algorithms

While the exact N-Best algorithm is efficient, the computation required to produce N hypotheses is roughly proportional to N. Typically, we needed to produce around 20 hypotheses to ensure that either the correct sentence is included, or the natural language components are sure to find an acceptable sentence. In an effort to reduce this computation, we have devised an algorithm called the Word-Dependent N-Best search algorithm. This algorithm merges the computation for different theories if the preceding word is the same. The result is that the number of alternative theories that must be computed is greatly reduced from 20 to between 3 and 6. Thus, the computation has been greatly reduced relative to the exact Sentence-Dependent Algorithm. We also implemented an algorithm that reduces the computation much further. The algorithm, called the Lattice N-Best Algorithm, requires no more work than that required for computation of the 1-Best sentence.

To compare the accuracy of the two approximate N-Best algorithms (Lattice and Word-Dependent) with that of the exact (Sentence-Dependent) N-Best algorithm, we computed the 100-Best sentences using all three algorithms. We found that the Word-Dependent Algorithm had an accuracy that was essentially equivalent to that of the Sentence-Dependent algorithm. Both of these algorithms found the correct sentence 96% of the time within the top 100 sentence hypotheses. In contrast, the Lattice algorithm found only 92% of the sentences, which means twice as many sentences were not found. Therefore, we have decided to use the Word-Dependent N-Best algorithm in our real-time spoken language system.

Natural Language Understanding

We have given the name Delphi to the natural language component of our SLS system, HARC. Delphi includes unification-based syntax and semantics, a parser, lexical and morphological components, and a discourse processor. Delphi is designed to minimize the number of incorrect answers to queries, as well as maximizing the number of correct answers.

Parser and Pre-Processor

We modified the pre-processor (that part of the system which modifies an input string into a form that the parser can process) to allow it to handle various written forms of time expressions (such as 1800 and 10:15 am), to facilitate the handling of synonyms, and to remove words such as "please" and "thank-you" which do not contribute to the semantic interpretation of the utterance.

We improved the performance of the parser by adding a facility for prediction. Formerly, the parser searched all possible assignments of syntactic structure for every sub-string of the input sentence, without taking into account the context of the sentence to the left of the sub-string. At every point in the sentence being parsed, the parser now uses the context established by the already-parsed portion of the sentence to predict what major phrasal categories could grammatically follow. Each partially-matched rule establishes expectations for the categories of items needed for it to continue, and only categories that are expected are searched for.

As a result of these an other changes, the average number of parses per sentence for the personnel corpus has dropped from 5 to 3 parses per sentence. We streamlined and simplified some aspects of the grammar and the parser, resulting in a speedup of more than a factor of 15 in parsing time, with further speedup still possible.

Lexicon, Syntax, and Semantics

The ATIS domain offered some interesting new syntactic and semantic phenomena, and we extended Delphi to parse and interpret these constructions correctly. One such phenomena is that the ATIS data exhibits greater freedom of expression in the ordering of constituents than is common in database retrieval tasks. This difference may be due to the ATIS task domain itself, or it may be due to the fact that the language was spoken rather than written, and thus may be more informal. To accommodate this type of language, we made changes primarily in the grammar and the lexicon.

We also relaxed some normal grammatical constraints such as subject-verb agreement to allow Delphi to be more tolerant of common deviations from standard grammar which are more common in spoken language than in written language.

We increased coverage of our integrated grammar (which includes syntax and semantics) on our personnel corpus of 761 sentences to 70%. We also added a treatment of temporal modification, because time is an important issue in a variety of domains, and is particularly crucial in the ATIS domain.

We developed a semantic treatment of time that allows for dependency on discourse context and on the tense of calendar references. For example, when processing a sentence like "I left on June 16" or "I will leave on June 16" it is necessary for the system to resolve "June 16" to a particular date, including the year, which must be different years in the two example sentences.

We also incorporated domain-independent solutions to certain common problems of semantic interpretation. These include handling the different ways in which a relationship of possession can be expressed in English ("the cost of the trip" and "the trip's cost", for example), and giving a common treatment to temporal modifiers occurring in both clausal and noun phrase contexts ("a November trip", "a trip in November", "...will leave in November").

Discourse Module

We added to our discourse module several new capabilities. Among them was a facility for head-noun and noun-phrase ellipses. An example of the latter is the dialogue consisting of two queries "What airlines fly to Washington?" and "Dallas?" Here the second question, though not a complete sentence, is understood to be a shorthand for "What airlines fly to Dallas?".

We also added an initial capability for handling definite references. Definite references are noun phrases such as "this person" and "the salaries" that are intended to refer to a specific entity or a group of entities. Our system now uses the semantic class information in the noun phrase to search for an entity in the preceding discourse that the definite phrase refers to. In the case of a definite reference that contains an open slot to be filled, ("the salary"), the system looks for an entity in the preceding discourse that can fill the slot.

TRANSCOM Demonstration Domain

After careful consideration of a number of promising domains, we selected the TRANSCOM domain as our application for our demonstration spoken language system. USTRANSCOM (TRANSportation COMmand) is responsible for planning the inter-theatre movement of personnel, material, and supplies around the world for the Army, Navy, Air Force, and other services. The DART project (Dynamic Analysis Replanning Tool) at BBN, sponsored by DARPA and RADC, is providing a quick demonstration of the operational impact of AI planning and scheduling technology on transportation planning at USTRANSCOM.

This application is very real, and very military. If it is necessary to have "real" users, they can be found as close as Scott Air Force Base near St. Louis. Despite the military nature of the application, the general concept of planning movements of people and supplies is meaningful to people without knowledge of military operations, and thus the demonstrations we develop will be understanable by non-military people. The development database is in Oracle, is unclassified, and is currently running on a Sun at BBN.

We have outlined a number of levels of potential demonstrations, and have begun develop specifications for the first demonstration, which we expect to be ready by February.

Speech Recognition

Detecting New Vocabulary Words

Our initial experiments for detecting new vocabulary words have been completed. We used an explicit, but general model for the acoustics of arbitrary words to recognize the existence of new words. The model allows for any sequence of phonemes at least two phonemes long. This general word model is then defined as being a member of each of the open class categories in the statistical grammar. We ran experiments on 175 sentences, spoken by a total of 7 speakers. A total of 62 of the words in the test sentences were not in the dictionary and grammar. The algorithm detected 71% of the missing words correctly, while spuriously detecting new words (false alarms) in only 0.6% of the test sentences. This false alarm rate is low enough that the algorithm could be included in a real system without fear of annoying the user.

Adding New Vocabulary Words

Now that we have demonstrated a basic capability to detect when the user speaks a word that is outside the vocabulary, we are developing techniques for adding the new word to the vocabulary. We assume that the user will be asked to type the word, since this is the only way that the system can be sure that it is, in fact, a new word. The system then must be able to create an acoustic model for the new word so that when it is spoken again it will be able to recognize the word.

In our current approach, we first look for the potentially new word in a large (150,000) entry phonetic lexicon. If the word is not in the lexicon we create a phonetic spelling by giving the orthographic spelling to the DECTalk synthesizer. Since DECTalk often produces errors in the phonetic spelling we also use a spoken sample of the word to determine the phonetic spelling. Based on the spelling produced DECTalk we create a network of likely phonetic spellings. Then we ask the user to speak the word. We use the network as a grammar that will constrain the phonetic recognition so that the final spelling is determined from both sources of knowledge.

We have implemented the use of the large dictionary and the connection between BYBLOS and DECTalk. Initial experiments indicate that most words are found in the large dictionary. About 20% of the phonetic spellings produced by DECTalk are incorrect. When these incorrect spellings are used the word error rate for these words increases somewhat. We have not yet performed the constrained phonetic recognition experiments.

Real-Time Spoken Language System

One of the requirements of this contract is to demonstrate a real-time spoken language system. We have been following the various efforts in the program for producing suitable hardware, but we have become concerned that those efforts will not result in an acceptable, timely solution. The VLSI efforts at SRI and Berkeley have not yet resulted in any working hardware. The PLUS hardware being developed at CMU is based on the relatively slow Motorola 88000 chip, and requires that the recognition be implemented in parallel on several boards. Therefore we decided that it would be much easier and safer to use commercially available, general purpose computing boards with large amounts of computing power and fast memory.

Both Sky Computer and Mercury produce a board based on the Intel 860 chip, which is about three times as fast as the Motorola 88000, and thus may provide enough speed so that recognition can be accomplished using only one or two boards. In addition, these boards both offer C compilers, so that machine-independent programs that were developed on the SUN can simply be recompiled and ported.

In order to make real-time recognition feasible we derived and implemented several new algorithms to speed up the recognition search for the N Best sentences. These algorithms are described below.

Fast Search with Statistical Grammars

One of the requirements that we place on the system is that it have a robust grammar that allows the user to speak naturally. Word-pair grammars are not acceptable because they greatly restrict the allowable sentences. Therefore, we use a fully connected statistical grammar based on pairs of classes of words. However, since all word classes are allowed to follow each word, the grammar computation will grow as the square of the number of word classes, and this grammar computation tends to dominate the total computation. We developed an algorithm that reduces the computation needed for fully-connected statistical grammars by a factor of 5 to 20, depending on the size of the original grammar.

Forward-Backward Search

The most effective way to achieve speedup is to avoid computation for sequences that are unlikely. We devised a multiple-pass search strategy that we call the Forward-Backward Search Algorithm, which uses a simplified forward pass on the whole sentence to derive information that can be used to speed up a detailed backward second pass search by a large factor. The backward pass is sped up because it can use the forward pass scores to predict which of the many theories will result in high scores. For our real-time effort, we use a 1-Best search forward as the speech is coming in to the system, and then perform the much more expensive N-Best search in the backwards direction.

The result is that the N-Best search computation is reduced by a factor of 40 *with no increased search errors!*

Summary of Speed Improvements

We have done several things to speed up the speech recognition computation of the N best sentences. The new algorithms include the new Word-Dependent N-Best algorithm, the technique for reducing statistical grammar computation, and the Forward-Backward Search. In addition, we sped up the code through careful implementation, and we expect a factor of 4 to 5 in speed by using the Intel 860 boards. The following table enumerates the methods and their speed improvements.

Method	Speedup Factor
Statistical grammar algorithm	5
Word-Dependent N-Best	5
Forward-Backward Search	40
Code Optimization	4
Intel 860 Board	4
Total reduction in computation	**16,000**

As a result of these improvements, the time necessary to compute the N-Best sentences has been reduced from about 10,000 times real-time to about 1/2 times real-time!

This computation will take place after the sentence has been spoken, but in much less than real-time, so the delay will be quite short. In fact, the computation of the N-Best sentences will happen during the same time that the natural language component is parsing the 1-Best sentence hypothesis, so the delays are not additive in the system as a whole.

Real-Time Demonstration

We implemented a real-time demonstration of speech recognition that produces the top N sentences. This required implementing a complete front end that would filter, sample, analyze, and vector quantize the speech, and pass the results on to the recognition search. We used a programmable MTU filter and A/D converter to do the basic speech sampling. We used a Sky Challenger with two TMS320C30 processors to control the MTU and to perform the signal processing (Mel-Frequency Cepstral Analysis) and vector quantization. The Sky Challenger was placed on the VME bus of a SUN 4/330 workstation.

While our plan was to use the SkyBolt signal processing board for the recognition search, we found that, since we had sped up the recognition so much, the recognition was able to run in almost real-time on the SUN 4 workstation without the need for further accelerator boards!

All of this was implemented in a demonstration. The signal processing and vector quantization is performed in real-time, while the speaker is speaking. The forward pass recognition search also takes place at the same time. Shortly after the speaking stops, the system finishes the recognition of the most likely sentence, and prints the answer onto the screen. It plays the speech back to the user so that s/he can verify that the answer is correct. (In a spoken language system, this answer will be fed to the natural language component for understanding.) Meanwhile, the system performs a backward search for the N-Best sentences, which are displayed on the screen with their corresponding acoustic and statistical language model scores. The backward pass is fast enough so that it is always finished long before the sentence has been replayed to the speaker. (In the spoken language system, these N answers will be made available to the natural language component, in case the first choice sentence did not parse or did not make sense.)

Speaker-Independent Demonstration

We created a speaker-independent speech model using the speech of eight male speakers, using the new speaker-independent training paradigm that we developed as part of our basic research contract in continuous speech recognition. We repeated the same steps for the seven females available. The demonstration used a statistical first-order class grammar that allows all sequences of words with some probability. This demonstration has now been shown to several government visitors.

During the next year we plan to connect the speech recognition system to the natural language component to produce a near-real-time spoken language demonstration. This will require collecting some speech from the new domain in order to create phonetic models for the vocabulary of that domain.

New Batch Queueing System

We have developed a mechanism that allows several users to submit batch jobs to a central queue, which then automatically runs these jobs on several compute servers. The mechanism is fairly general in that it allows simultaneous control of job queues on different types of machines and from different projects. It is also more robust than the standard UNIX batch mechanism, in that it keeps better track of running jobs. This mechanism makes it quite feasible to use a large number of workstations efficiently for research computing. Each researcher submits a sequence of jobs to their "preferred machine". However, jobs will also run on other machines that are idle. Thus all the machines are used almost all the time. This mechanism would make it feasible to obtain a large amount of computing resources while still taking advantage of the lower cost of midrange computing.

Principal Investigator: John Makhoul
PI Institution: BBN Systems and Technologies
PI Phone Number: 617-873-3332
PI E-mail Address: makhoul@bbn.com
Grant or Contract Title: Spoken Language Systems
Grant or Contract Number: N00014-89-C-0008
Reporting Period: 1 Oct 89 - 30 Sep 90

3. Lists of publications, presentations, reports, and awards/honors.

A. Refereed papers/presentations published:

A. Asadi, R. Schwartz, and J. Makhoul, *Automatic Detection of New Words in a Large Vocabulary Continuous Speech Recognition System*, IEEE International Conference on ASSP, Albuqueruqe, NM, April, 1990.

A. Asadi, R. Schwartz, J. Makhoul, *Automatic Detection of New Words in a Large Vocabulary Continuous Speech Recognition System*, DARPA Speech and NL Workshop, Cape Cod, Mass., October 1989.

S. Austin, P. Peterson, P. Placeway, R. Schwartz, and J. Vandegrift, *Toward a Real-Time Commerical System Using Commercial Hardware*, DARPA Speech and NL Workshop, Hidden Valley, PA, June, 1990.

M. Bates, S. Boisen, J. Makhoul, *Developing an Evaluation Methodology for Spoken Language Systems*, DARPA Speech and NL Workshop, Hidden Valley, PA, June, 1990.

M. Bates, R. Bobrow, S. Boisen, R. Ingria, and D. Stallard, *BBN ATIS Progress Report - June 1990*, DARPA Speech and NL Workshop, Hidden Valley, PA, June, 1990.

R. Bobrow, R. Ingira, and D. Stallard, *Syntactic and Semantic Knowledge in the DELPHI Unification Grammar*, DARPA Speech and NL Workshop, Hidden Valley, PA, June, 1990.

R. Bobrow and L. Ramshaw, *On Deftly Ingtroducing Procedural Elements into Unification Parsing*, DARPA Speech and NL Workshop, Hidden Valley, PA, June, 1990.

S. Boisen, L. Ramshaw, D. Ayuso, M. Bates, *A Proposal for SLS Evaluation*, DARPA Speech and NL Workshop, Cape Cod, Mass., October 1989.

Y-L. Chow, *Maximum Mutual Information Estimation of HMM Parameters for Continuous Speech Recognition Using the N-Best Algorithms*, IEEE International Conference on ASSP, Albuqueruqe, NM, April, 1990.

Y-L. Chow and R. Schwartz, *The N-Best Algorithm: An Efficient Procedure for Finding the Top N Sentence Hypotheses*, DARPA Speech and NL Workshop, Cape Cod, Mass., October 1989.

A. Derr, R. Schwartz, *A Simple Statistical Class Grammar for Measuring Speech Recognition Performance*, DARPA Speech and NL Workshop, Cape Cod, Mass., October 1989.

R. Ingria, *The Limits of Unification*, 28th Annual Meeting of the Association for Computational Linguistics, Morristown, NY, June, 1990.

R. Ingria, L. Ramshaw, *Porting to New Domains Using the Learner,* DARPA Speech and NL Workshop, Cape Cod, Mass., October 1989.

R. Schwartz and S. Austin, *Efficient, High-Perofrmance Algorithms for N-Best Search*, DARPA Speech and NL Workshop, Hidden Valley, PA, June, 1990.

R. Schwartz and Y-L. Chow, *The N-Best Algorithm: An Efficient and Exact Procedure for Finding the N Most Likely Sentence Hypotheses*, IEEE International Conference on ASSP, Albuqueruqe, NM, April, 1990.

D. Stallard, *Unification-Based Semantic Interpretation in the BBN Spoken Language System*, DARPA Speech and NL Workshop, Cape Cod, Mass., October 1989.

B. Books or sections thereof:

R. Ingria, *Simulation of Language Understanding: Lexical Recognition*, in Computational Linguistics: An International Handbook on Computer Oriented Language Research and Applicatons, Walter de Gruyter, Berlin, New York, 1990, pp. 336-347.

(submitted, not yet published) R. Ingria and Leland Maurice George, *Adjectives, Nominals, and the Status of Arguments*, in James Pustejovsky, ed., Semantics and the Lexicon, Kluwer Academic Publishers, Dordrecht, the Netherlands, to appear 1991.

C. Invited presentations:

M. Bates and R. Weischedel, "Challenging Problems for Natural Language Research", presented at the Workshop on Future Directions in Natural Language Processing, BBN, Cambridge, MA, December, 1989.

R. Ingria, ``Grammar Engineering in Delphi", talk presented at the Grammar Engineering Workshop, University of Saarbrucken, June 22, 1990.

R. Ingria, ``Grammar Development and Evaluation in the BBN Spoken Language System", talk presented at the University of Chicago Center for Information and Language Studies, Chicago, May 21, 1990.

D. Other presentations: none

E. Unrefereed papers:

R. Ingria and J. Pustejovsky, *Active Objects in Syntax, Semantics, and Parsing'*, in Carol Tenny, ed., Papers from the Parsing Seminar, MIT Center for Cognitive Science, 1990

Principal Investigator: John Makhoul
PI Institution: BBN Systems and Technologies
PI Phone Number: 617-873-3332
PI E-mail Address: makhoul@bbn.com
Grant or Contract Title: Spoken Language Systems
Grant or Contract Number: N00014-89-C-0008
Reporting Period: 1 Oct 89 - 30 Sep 90

4. Transitions and DoD interactions.

Since all of the work we do is presented at regular DARPA workshops in which all contractors, as well as outside organizatioins, are present, the technical work we do gets transferred immediately to those people attending. Organizations represented at those meetings include universities, national laboratories, industry, and government agencies. In addition, the proceedings of these workshops are distributed widely and are sold by a publisher. These workshops provide a frequent forum for interaction with DoD agency represnetatives. In addition, every year we present our work at major confrences such as the IEEE International Conference on Acoustics, Speech, and Signal Processing, and the Association for Computational Linguistics.

Principal Investigator: John Makhoul
PI Institution: BBN Systems and Technologies
PI Phone Number: 617-873-3332
PI E-mail Address: makhoul@bbn.com
Grant or Contract Title: Spoken Language Systems
Grant or Contract Number: N00014-89-C-0008
Reporting Period: 1 Oct 89 - 30 Sep 90

5. Software and hardware prototypes.

Software: We are continuing to devel the Delphi natural language software and the rest of the software that comprises the HARC system.

Hardware: None.